Applied Microeconometrics Problem Set 2

Tanya Rajan

December 1, 2020

Question 1. Consider the nonparametric Roy model

$$
\begin{aligned}
Y &= DY(1) + (1-D)Y(0) \\
D &= \mathbb{1}[U \leq p(Z)]
\end{aligned}
$$

where $(Y(0), Y(1), U)$ are unobserved and assumed to be independent of Z, and $p(z) \equiv \mathbb{P}[D = 1 \mid Z = z]$ is the propensity score. Suppose that U is continuously distributed and that it has been normalized to have a uniform distribution over $[0, 1]$. Suppose that $\mathbb{E}\left[Y(d)^2\right]$ exists for both $d = 0, 1$. Define the marginal treatment effect (MTE) as $m(u) \equiv \mathbb{E}[Y(1) - Y(0) \mid U = u]$ for any $u \in [0, 1]$

(a) Show that the average treatment on the untreated (ATU) can be written as the following weighted average of the MTE function:

$$
\text{ATU} \equiv \mathbb{E}[Y(1) - Y(0) \mid D = 0] = \int_0^1 m(u) \frac{\mathbb{P}[p(Z) < u]}{\mathbb{P}[D = 0]} du
$$

$$
\begin{aligned}
\mathbb{E}[Y(1) - Y(0) \mid D = 0] &= \mathbb{E}[Y(1) - Y(0) \mid U > p(z)] && \text{using def. of } D \\
&= \mathbb{E}[\mathbb{E}[Y(1) - Y(0) | U > p(z), U = u] \mid U > p(z)] && \text{by LIE} \\
&= \mathbb{E}[\underbrace{\mathbb{E}[Y(1) - Y(0) \mid U = u}_{=MTE(u)}] \mid U > p(z)] && \text{by independence} \\
&= \frac{\mathbb{E}[MTE(u)\mathbb{1}\{U > p(z)\}]}{\mathbb{P}(U > p(z))} && \text{def. of cond'l. prob.} \\
&= \frac{\mathbb{E}\left[\mathbb{E}[MTE(u)\mathbb{1}\{U > p(z)\} \mid U = u]\right]}{\mathbb{P}(D = 0)} && \text{by LIE} \\
&= \frac{\mathbb{E}[MTE(u)\mathbb{P}(u > p(z))]}{\mathbb{P}(D = 0)} && \text{evaluating inner expectation} \\
&= \int_0^1 m(u) \frac{\mathbb{P}(u > p(z))}{\mathbb{P}(D = 0)} du
\end{aligned}
$$

(b) Show that the Wald estimand from a contrast between two instrument values z_0 and z_1 with $p(z_1) > p(z_0)$ can be written as the following weighted average of the MTE function:

$$\frac{\mathbb{E}[Y \mid Z = z_1] - \mathbb{E}[Y \mid Z = z_0]}{\mathbb{E}[D \mid Z = z_1] - \mathbb{E}[D \mid Z = z_0]} = \int_0^1 m(u) \left(\frac{\mathbb{1}[p(z_0) < u \leq p(z_1)]}{p(z_1) - p(z_0)} \right) du$$

Explain the connection with what we know about the local average treatment effect (LATE)

Let us define $D(z_1) = \mathbb{1}\{U \leq p(z_1)\}$, and similarly for $D(z_0)$. Note that since $U \perp Z$, it must be that $D(z_1) \perp Z$. Then we can write:

$$\begin{aligned}
\mathbb{E}[Y \mid Z = z_1] - \mathbb{E}[Y \mid Z = z_0] &= \mathbb{E}[Y(0) + D(z_1)(Y(1) - Y(0)) \mid Z = z_1] \\
&\quad - \mathbb{E}[Y(0) + D(z_0)(Y(1) - Y(0)) \mid Z = z_0] \\
&= \mathbb{E}[Y(0) + D(z_1)(Y(1) - Y(0))] - \mathbb{E}[Y(0) + D(z_0)(Y(1) - Y(0))] \\
&= \mathbb{E}[(D(z_1) - D(z_0))(Y(1) - Y(0))] \\
&= \mathbb{E}[Y(1) - Y(0) | D(z_1) > D(z_0)] \mathbb{P}(D(z_1) > D(z_0)) \\
&\quad - \mathbb{E}[Y(1) - Y(0) | D(z_1) > D(z_0)] \mathbb{P}(D(z_1) < D(z_0))
\end{aligned}$$

where the first line follows from the definition of Y, the second line follows from independence, the third from rearranging the terms in the expectations, and the fourth line follows from the law of total probability (when $D(z_1) = D(z_0)$, the expectation evaluates to zero). Here we note that since we are given $p(z_1) > p(z_0)$, it must mean that $D(z_1) > D(z_0)$ from the definition of $D(z_d)$. So we can rule out the second term in the final equation above. Further, note that

$$\mathbb{P}(D(z_1) > D(z_0)) = \mathbb{P}(\mathbb{1}\{U \leq p(z_1)\} > \mathbb{1}\{U \leq p(z_0)\}) = \mathbb{P}(p(z_0) < U \leq p(z_1))$$

As a result, we can write the numerator of the Wald estimand as:

$$\mathbb{E}[Y \mid z = z_1] - \mathbb{E}[Y \mid z = z_0] = \mathbb{E}[Y(1) - Y(0) | p(z_0) < U \leq p(z_1)] \mathbb{P}(p(z_0) < U \leq p(z_1))$$

Looking at the denominator of the Wald estimand, we can use similar steps as above to get:

$$\begin{aligned}
\mathbb{E}[D \mid Z = z_0] - \mathbb{E}[D \mid Z = z_0] &= \mathbb{E}[D(z_1) \mid Z = z_1] - \mathbb{E}[D(z_0) \mid Z = z_0] \\
&= \mathbb{E}[D(z_1) - D(z_0)] \\
&= \mathbb{P}(D(z_1) > D(z_0)) \\
&= \mathbb{P}(\mathbb{1}\{U \leq p(z_1)\} > \mathbb{1}\{U \leq p(z_0)\}) \\
&= \mathbb{P}(p(z_0) < U \leq p(z_1))
\end{aligned}$$

The Wald estimand is then just $\mathbb{E}[Y(1) - Y(0) | p(z_0) < U \leq p(z_1)]$. We can expand this using the definition of the conditional expectation and then apply LIE to get the desired result. Note that $\mathbb{P}(p(z_0) < U \leq p(z_1)) = p(z_1) - p(z_0)$ since $U \sim Unif[0, 1]$.

$$\begin{aligned}
\frac{\mathbb{E}[Y(1) - Y(0) \mathbb{1}\{p(z_0) < U \leq p(z_1)\}]}{\mathbb{P}(p(z_0) < U \leq p(z_1))} &= \frac{\mathbb{E}[Y(1) - Y(0) \mathbb{1}\{p(z_0) < U \leq p(z_1)\}]}{\mathbb{P}(p(z_0) < U \leq p(z_1))} \\
&= \frac{\mathbb{E}[\mathbb{E}[Y(1) - Y(0) \mathbb{1}\{p(z_0) < U \leq p(z_1)\} \mid U = u]]}{p(z_1) - p(z_0)} \\
&= \frac{\mathbb{E}[MTE(u) \mathbb{1}\{p(z_0) < u \leq p(z_1)\}]}{p(z_1) - p(z_0)} \\
&= \int_0^1 m(u) \left(\frac{\mathbb{1}[p(z_0) < u \leq p(z_1)]}{p(z_1) - p(z_0)} \right) du
\end{aligned}$$

Using this formulation, we can see that the LATE is identifying sections of the MTE curve. For example, for points z_1 and z_0, it is identifying the MTE on the support of U between $p(z_0)$ and $p(z_1)$. If we selected two different instruments (or instrument values) that satisfied the necessary assumptions, we may get an estimate of a different section of the MTE curve.

(c) Let $D^\star \equiv \mathbb{1}\left[U \leq p^\star\left(Z^\star\right)\right]$ and $Y^\star \equiv D^\star Y(1)+\left(1-D^\star\right) Y(0)$ for some function $p^\star$ where $Z^\star(Y(0), Y(1), U)$. Show that the policy-relevant treatment effect (PRTE) can be written as the following weighted average of the MTE function:

$$\mathrm{PRTE} \equiv \frac{\mathbb{E}\left[Y^\star\right]-\mathbb{E}[Y]}{\mathbb{E}\left[D^\star\right]-\mathbb{E}[D]}=\int_0^1 m(u)\left(\frac{F_P^{-}(u)-F_{P^\star}^{-}(u)}{\mathbb{E}\left[p^\star\left(Z^\star\right)\right]-\mathbb{E}[p(Z)]}\right) du$$

where $F_P^{-}(u) \equiv \lim_{v \uparrow u} F_P(v)$, and similarly for $F_{P^\star}^{-}(u)$

I follow the steps outlined in Appendix C of the Heckman and Vytlacil (2007) handbook chapter II.[1] First, let us notationally define $\mathbb{1}_S(a)$ to mean that the indicator is turned on when $a \in S$. Considering only the non-starred policy, we can write:

$$\begin{aligned}
\mathbb{E}[Y] &= \int_0^1 \mathbb{E}[Y \mid p(Z)=t] dF_P(t) \\
&= \int_0^1\left[\int_0^1\left(\underbrace{\mathbb{1}_{[0,t]}(u) \mathbb{E}[Y(1) \mid U=u]}_{\text{assoc. w}/U \leq p(Z)=t}+\underbrace{\mathbb{1}_{(t,1]}(u) \mathbb{E}[Y(0) \mid U=u]}_{\text{assoc. w}/U>p(Z)=t}\right) du\right] dF_P(t)
\end{aligned}$$

Now note that we are given that $\mathbb{E}[Y(d)^2]<\infty$, which means that $\mathbb{E}[Y(d)]<\infty$ as well. Further, we can write:

$$\mathbb{E}[|\mathbb{1}_{[0,t]}(u) \mathbb{E}[Y(1) \mid U=u]+\mathbb{1}_{(t,1]}(u) \mathbb{E}[Y(0) \mid U=u]|] \leq \mathbb{E}[|Y(1)|+|Y(0)|]<\infty$$

Given this condition, we can use Fubini's theorem to switch the order of integration.

$$\begin{aligned}
\mathbb{E}[Y] &= \int_0^1\left[\int_0^1\left(\mathbb{1}_{[u,1]}(u) \mathbb{E}[Y(1) \mid U=u]+\mathbb{1}_{[0,u)}(u) \mathbb{E}[Y(0) \mid U=u]\right) dF_P(t)\right] du \\
&= \int_0^1\Big((1-F_P^{-}(u)) \mathbb{E}[Y(1) | U=u]+F_P^{-}(u) \mathbb{E}[Y(0) | U=u]\Big) du
\end{aligned}$$

We can do a similar exercise for $E[Y^*]$ where everything is the same except the distribution is now F_{P^*} and the propensity score is $p^*(z)$. Then considering the numerator of the PRTE, we have

$$\begin{aligned}
\mathbb{E}\left[Y^\star\right]-\mathbb{E}[Y] &= \int_0^1\left(\mathbb{E}[Y(1) \mid U=u]\left(1-F_{P^*}^{-}(u)-1+F_P^{-}(u)\right)-\mathbb{E}[Y(0) \mid U=u]\left(F_{P^*}^{-}(u)-F_P^{-}(u)\right)\right) du \\
&= \int_0^1 \underbrace{\mathbb{E}[Y(1)-Y(0) \mid U=u]}_{m(u)}\left(F_P^{-}(u)-F_{P^*}^{-}(u)\right) du
\end{aligned}$$

Considering the denominator of the PRTE, we can say:

$$\mathbb{E}[D]=\mathbb{E}[\mathbb{E}[D | Z]]=\mathbb{E}[\mathbb{P}(D=1 | Z)]$$

[1] Heckman, J.J., and Vytlacil, E.J. (2007) Econometric evaluation of social programs part II: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. Volume 6, Part B.

where the last equality follows from the law of total probability and the second term associated with $D = 0$ evaluates to 0, so disappears from the final equation. Then we have the desired denominator $\mathbb{E}[D^*] - \mathbb{E}[D] = \mathbb{E}[p^*(Z^*)] - \mathbb{E}[p(Z)]$. Putting it all together, we have

$$\int_0^1 m(u) \left(\frac{F_P^-(u) - F_{P^*}^-(u)}{\mathbb{E}\left[p^*\left(Z^*\right)\right] - \mathbb{E}[p(Z)]} \right) du$$

(d) Define the marginal treatment response (MTR) function as $m(d \mid u) \equiv \mathbb{E}[Y(d) \mid U = u]$. Let s be a function of D and Z. Show that

$$\begin{aligned} \mathbb{E}[s(D, Z)Y] = \textstyle\int_0^1 \quad & m(1 \mid u) \times \mathbb{E}[s(1, Z) \mid P \geq u] \left(1 - F_P^-(u)\right) du \\ & + \textstyle\int_0^1 m(0 \mid u) \times \mathbb{E}[s(0, Z) \mid P < u] F_P^-(u) du \end{aligned}$$

where (as usual) $P \equiv p(Z)$ is the propensity score viewed as a random variable. Minor mathematical note: If $\mathbb{P}[P \geq u] = 0$ or $\mathbb{P}[P < u] = 0$ then, strictly speaking, the conditional expectations are not well-defined. However, in either case they are multiplied by 0, so we can safely ignore this little detail.

We can start out by taking the same steps as we did in part (c) above.

$$\begin{aligned} \mathbb{E}[\mathrm{s}(D, Z)Y] &= \int_0^1 \mathbb{E}[s(D, Z)Y \mid p(Z) = t] dF_P(t) \\ &= \int_0^1 \left[\int_0^1 \Big(\mathbb{1}_{[0,t]}(u) \mathbb{E}[s(1, Z) | P = t] \mathbb{E}[Y(1) \mid U = u] \right. \\ &\qquad \left. + \mathbb{1}_{(t,1]}(u) \mathbb{E}[s(0, Z) | P = t] \mathbb{E}[Y(0) \mid U = u] \Big) du \right] dF_P(t) \\ &= \int_0^1 \left(\int_0^1 \left(\mathbb{1}_{[u,1]} \mathbb{E}[s(1, Z) | P = t] \mathbb{E}[Y(1) | U = u] + \mathbb{1}_{[0,u)} \mathbb{E}[s(0, Z) | P = t] \mathbb{E}[Y(0) | U = u] \right) dF_P(t) \right) du \\ &= \int_0^1 \left[\int_0^1 \left(\mathbb{1}_{[u,1)}(u) \mathbb{E}[s(1, Z) | P \geq u] m(1|u) + \mathbb{1}_{[0,u)}(u) \mathbb{E}[s(0, Z) | P < u] m(0|u) \mid U = u] \right) du \right] dF_P(t) \\ &= \int_0^1 m(1|u) \mathbb{E}[s(1, Z) | P \geq u] (1 - F_P(u)) du + \int_0^1 m(0|u) \mathbb{E}[s(0, Z) | P < u] F_P(u) du \end{aligned}$$

(e) Construct an example under which the two-stage least squares (TSLS) estimand is a weighted average of $m(u)$ with weights that are negative for some values of u.

I follow Mogstad, Torgovistky, Walter, (2018). If we have multiple instruments that are not ordered in an appropriate way, we get that there are compliers for each instrument. But then the sign of the weight will be determined by:

$$\operatorname{sign}\left(\omega_{2c} = \mathbf{1}\left[\pi_{2,\text{ complier }\mathbf{s}} > 0\right] \operatorname{sign}\left[P\left(D_i = 1 \mid Z_{i,2} = 1\right) - P\left(D_i = 1 \mid Z_{i2} = 0\right)\right]\right.$$

This can give us negative weights.

Question 2. Suppose that $D \in \{0,1\}$ is a binary treatment and let $(Y(0), Y(1))$ be the potential outcomes associated with D. Assume that D is determined by

$$D = \mathbb{1}[U \leq p(Z)]$$

where $p(z) \equiv \mathbb{P}[D = 1 \mid Z = z]$ is the propensity score, U is an unobservable random variable that is distributed uniformly over $[0,1]$, and $Z \in \{0,1\}$ is a binary instrument. Assume that $(Y(0), Y(1), U) \perp Z$, and that $p(1) > p(0)$

(a) Suppose that we also assume that the marginal treatment response functions are linear, i.e.

$$\mathbb{E}[Y(d) \mid U = u] = \alpha_d + \beta_d u \quad \text{for } d = 0,1$$

Show that there is no unobserved heterogeneity in the causal effect of D on Y (i.e. the MTE function is a constant function of u) if and only if

$$\begin{aligned}&\mathbb{E}[Y \mid D = 1, Z = 1] - \mathbb{E}[Y \mid D = 1, Z = 0] \\ &\quad = \mathbb{E}[Y \mid D = 0, Z = 1] - \mathbb{E}[Y \mid D = 0, Z = 0]\end{aligned}$$

Let us first work with the expression: $\mathbb{E}[Y|D = 1, Z = 1] = \mathbb{E}[Y|D = 1, Z = 0]$. We can say this is equal to:

$$\begin{aligned}\mathbb{E}[Y \mid D = 1, Z = 1] - \mathbb{E}[Y, D = 1, Z = 0] &= \mathbb{E}[Y(1) \mid D = 1, Z = 1] - \mathbb{E}[Y(1) \mid D = 1, Z = 0] \\ &= \mathbb{E}[Y(1) \mid U \leq p(1)] - \mathbb{E}[Y(1)|U \leq p(0)] \\ &= \mathbb{E}[\mathbb{E}[Y(1)|U = u] \mid U \leq p(1)] - \mathbb{E}[\mathbb{E}[Y(1)|U = u] \mid U \leq p(0)] \\ &= \mathbb{E}[\alpha_1 + \beta_1 U \mid U \leq p(1)] - \mathbb{E}[-\alpha_1 - \beta_1 U \mid U \leq p(0)] \\ &= \alpha_1 + \beta_1 \mathbb{E}\left[U \mid U \leq p(1)\right] - \alpha_1 - \beta_1 \mathbb{E}[U \mid U \leq p(0)] \\ &= \beta_1 \frac{p(1)}{2} - \beta_1 \frac{p(0)}{2} \\ &= \beta_1 \frac{(p(1) - p(0))}{2}\end{aligned}$$

where the second to last line follows from taking the conditional expectation of a uniform random variable on $[0,1]$. Similarly, we can say that $\mathbb{E}[Y|D = 0, Z = 1] = \mathbb{E}[Y|D = 0, Z = 0]$ is:

$$\begin{aligned}\mathbb{E}[Y \mid D = 0, Z = 1] - \mathbb{E}[Y, D = 0, Z = 0] &= \mathbb{E}[Y(0) \mid D = 0, Z = 1] - \mathbb{E}[Y(0) \mid D = 0, Z = 0] \\ &= \mathbb{E}[Y(0) \mid U > p(1)] - \mathbb{E}[Y(0)|U > p(0)] \\ &= \alpha_0 + \beta_0 \mathbb{E}\left[U \mid U > p(1)\right] - \alpha_0 - \beta_0 \mathbb{E}[U \mid U > p(0)] \\ &= \beta_0 \frac{1 + p(1)}{2} - \beta_0 \frac{1 + p(0)}{2} \\ &= \beta_0 \frac{(p(1) - p(0))}{2}\end{aligned}$$

Then it is the case that $\mathbb{E}[Y \mid D = 1, z = 1] - \mathbb{E}[Y, D = 1, Z = 0] = \mathbb{E}[Y \mid D = 0, z = 1] - \mathbb{E}[Y, D = 0, Z = 0] \iff \beta_1 = \beta_0$. Evaluating the MTE, we have:

$$\begin{aligned}\mathbb{E}[Y(1) - Y(0) \mid U = u] &= \mathbb{E}[Y(1) \mid U = u] - \mathbb{E}[Y(0) \mid U = u] \\ &= \alpha_1 + \beta_1 u - \alpha_0 - \beta_0 u \\ &= (\alpha_1 - \alpha_0) + (\beta_1 - \beta_0)\, u\end{aligned}$$

When $\beta_1 = \beta_0$, then the second term disappears and the MTE is a constant function of u as desired. The argument works in the opposite direction by the same logic.

(b) Suppose that we continue to assume that the MTR functions are linear as in (a). We regress Y on $p(Z)$ for each of the subpopulations $d = 0, 1$, to point identify α_d and β_d for $d = 0, 1$, and we use these quantities to construct an implied LATE using the interpretation of the LATE as a weighted average of the MTE functions. Show that this implied LATE is always equal to the Wald estimand, even if our assumption that the MTR functions are linear is not actually correct.

The LATE can be written as:

$$\begin{aligned}
\mathbb{E}\left[Y(1)-Y(0) \mid D\left(z_1\right)>D\left(z_0\right)\right] &= \mathbb{E}[Y(1)-Y(0) \mid D(1)=1, D(0)=0] \\
&= \mathbb{E}[Y(1)-Y(0) \mid p(0)<U<p(1)] \\
&= \mathbb{E}[Y(1) \mid p(0)<U<p(1)]-\mathbb{E}[Y(0) \mid p(0)<U<p(1)] \\
&= \frac{\mathbb{E}[Y(1) \mathbb{1}\{p(0)<U<p(1)\}]}{\mathbb{P}(p(0)<U<p(1))}-\frac{\mathbb{E}[Y(0) \mathbb{1}\{p(0)<U<p(1)\}]}{\mathbb{P}(p(0)<U<p(1))} \\
&= \int_{p(0)}^{p(1)} \frac{\mathbb{E}[Y(1) | U=u]}{p(1)-p(0)} du-\int_{p(0)}^{p(1)} \frac{\mathbb{E}[Y(0) | U=u]}{p(1)-p(0)} du \\
&= \int_{p(0)}^{p(1)} \frac{\alpha_1+\beta_1 u}{p(1)-p(0)} du-\int_{p(0)}^{p(1)} \frac{\alpha_0+\beta_0 u}{p(1)-p(0)} du
\end{aligned}$$

Evaluating these integrals using the uniform distribution for U, we get:

$$\begin{aligned}
&= \frac{1}{p(1)-p(0)}\left(\left(\alpha_1-\alpha_0\right) u+\frac{1}{2}\left(\beta_1-\beta_0\right) u^2\right)\Bigg|_{p(0)}^{p(1)} \\
&= \left(\alpha_1-\alpha_0\right)+\frac{1}{2}\left(\beta_1-\beta_0\right) \frac{p(1)^2-p(0)^2}{p(1)-p(0)} \\
&= \left(\alpha_1-\alpha_0\right)+\frac{1}{2}\left(\beta_1-\beta_0\right)(p(1)+p(0))
\end{aligned}$$

Note that when we regress Y on $p(Z)$ for each of the subpopulations $d = 0, 1$, we are working with a linear conditional expectation function (CEF) of the form $\mathbb{E}[Y|D = 1, p(Z)]$. Since Z is given to be a binary instrument in the problem, this CEF must be linear, which means that our regression recovers the exact CEF. So our linear estimate will equal the Wald estimand even if we are misspecified.

(c) Explain the significance of the finding in part (b) for using a linear MTR model.

We can use a linear MTR specification in the Wald estimand case with binary isntruments. Even if we are misspecified, we are able to say something about the data.

Question 3. Suppose that $D \in \{0, 1\}$ is a binary treatment, and Y is a continuously distributed outcome with potential outcomes $Y(0), Y(1)$. Let R be another continuously distributed observed random variable and suppose that $D = \mathbb{1}[R \geq c]$. Suppose there is an unobserved binary variable $M \in \{0, 1\}$ with the property that $\mathbb{P}[R \geq c \mid M = 1] = 1$. Let $f_{R|M}(r \mid m)$ denote the density of R conditional on M. Maintain the following assumptions:

- $f_{R|M}(r \mid 0)$ is continuous at $r = c$
- $f_{R|M}(r \mid 1)$ is continuous from the right at $r = c$, that is , $\lim_{r \downarrow c} f_{R|M}(r \mid 1) = f_{R|M}(c \mid 1)$
- $f_R(c) \neq 0$
- $\mathbb{E}[Y(0) \mid R = r, M = 0]$ is continuous at $r = c$
- The distribution of Y given r is continuous from the right at $r = c$, that is, $\lim_{r \downarrow c} \mathbb{P}[Y \leq y \mid R = r] = \mathbb{P}[Y \leq y \mid R = c]$ for all y

(a) Show that $\pi \equiv \mathbb{P}[M = 1 \mid R = c]$ is point identified

We are told that the variable M is only equal to 1 in the data above the cutoff. We want to identify π, the fraction of people with $M = 1$ exactly at the cutoff. Since we are given $f_{R|M}(r|M)$ is continuous at $r = c$, it must also be that $f_{R|M}(r|0)$ is continuous at $r = c$. If there is a jump s.t. $\lim_{r \uparrow c} f_{R|M}(r|0) \neq f(c|0)$, then it must be the case that the unobserved variable is causing the jump. So we can write:

$$\lim_{r \uparrow c} f_{R|M}(r|0) = f_{R|M}(c|0) = (1 - \pi) f_R(c)$$

Note this works because $1 - \pi = \mathbb{P}(M = 0|R = c)$. Then, we can rearrange to get

$$\pi = 1 - \frac{\lim_{r \uparrow c} f_{R|M}(r|0)}{f_R(c)}$$

Note that since we are given that $f_R(c) \neq 0$, π is cleanly identified.

(b) Derive sharp bounds on $\delta = \mathbb{E}[Y(1) - Y(0)|R = c, M = 0]$. *Hint: Apply a result from a previous problem set.*

First let us manipulate our given equation. We know that

$$\begin{aligned}
\delta &= \mathbb{E}[Y(1)|R = c, M = 0] - \mathbb{E}[Y(0)|R = c, M = 0] \\
&= \mathbb{E}[Y|R = c, M = 0] - \mathbb{E}[Y(0)|R = c, M = 0]
\end{aligned}$$

By assumption we know that $\mathbb{E}[Y(0)|R = r, M = 0]$ is continuous at $r = c$, and we know that $\mathbb{E}[Y|R = r] = \mathbb{E}[Y(0)|R = r, M = 0]$ for all $r < c$. So to identify the second term, we can take $\lim_{r \uparrow c} \mathbb{E}[Y|R = r] = \mathbb{E}[Y(0)|R = c, M = 0]$. Then we only need to consider the first term of δ above. Using the results from Problem Set 1, Question 2b, we can say that the first term is bounded by:

$$\mathbb{E}\left[Y \mid R = c, M = 0, Y \leq G^{-1}(1 - \pi)\right] \leq \mathbb{E}[Y|R = c, M = 0] \leq \mathbb{E}\left[Y \mid R = c, M = 0, Y \geq G^{-1}(\pi)\right]$$

where $G(\cdot)$ is the CDF of $Y|R = c, M = 0$ and π is the value we found in part (a) of this problem. So the full bound is:

$$\mathbb{E}\left[Y \mid R = c, M = 0, Y \leq G^{-1}(1 - \pi)\right] + \lim_{r \uparrow c} \mathbb{E}[Y|R = r] \leq \delta \leq \mathbb{E}\left[Y \mid R = c, M = 0, Y \geq G^{-1}(\pi)\right] + \lim_{r \uparrow c} \mathbb{E}[Y|R = r]$$

As I argued in problem set 1, question 2c, these bounds are sharp for distributions $Y(0), Y(1)$ where $Y(0) >> Y(1)$ or $Y(1) << Y(0)$. As in that question, I define $Y(0) >> Y(1)$ to mean that for all $y_1 \in \text{supp}(Y(1)), y_0 \in \text{supp}(Y(0))$, we have that $y_0 > y_1$. In these cases, we will have separation of the distributions and either the upper or lower bound will hold with equality.

> Explain how the assumptions and the result in part (b) differ from the usual sharp regression discontinuity framework.

The assumptions here are different than the usual sharp RD setting because we do not have the continuity of the running variable R on both sides of the cutoff. There is another unobserved variable, $M = 1$, that may be causing jumps around the cutoff, which may cause the running variable to appear discontinuous. Given the assumptions in the question on the conditioned running variable distribution, $f_{R|M}$, we can still bound the treatment effect around the cutoff by applying the arguments in parts (a) and (b) of this question.

Question 4. Let Y be an outcome, $D \in \{0,1\}$ a binary treatment, $Y(0), Y(1)$ potential outcomes, and R a running variable with interval support. Let $C \in \{c_\ell, c_h\}$ be a binary random variable, where c_ℓ, c_h are two known values in the support of R with $c_\ell < c_h$. Suppose that treatment is determined by

$$D = \mathbb{1}[R \geq C]$$

Assume that

$$\mathbb{E}[Y(d) \mid R = r, C = c] = g_d(r) + \alpha_d \mathbb{1}[c = c_h]$$

for all d, r, and c, where $g_d(r)$ is a continuous, unknown function of r, and α_d is an unknown scalar parameter. The observed data is (Y, D, R, C).
Show that $\mathbb{E}[Y(1) - Y(0) \mid R = r]$ is point identified for all $r \in [c_\ell, c_h]$. Provide an intuitive explanation of the role of the maintained assumptions in establishing this result and how both the assumptions and result differ from the usual identification argument for sharp regression discontinuity designs.

First let us work with the expectation we are given:

$$\begin{aligned}
\mathbb{E}[Y(1) - Y(0) \mid R = r] &= \mathbb{E}_c[\mathbb{E}[Y(1) - Y(0) \mid R = r, C = c]] \\
&= \mathbb{E}_c\left[g_1(r) + \alpha_1 \mathbb{1}\{C = c_h\} - g_0(r) - \alpha_0 \mathbb{1}\{C = c_h\}\right] \\
&= g_1(r) - g_0(r) + (\alpha_1 - \alpha_0)\,\mathbb{P}(C = c_h)
\end{aligned}$$

Define $\pi_h = \mathbb{P}(C = c_h)$. Since we observe C, this proportion is identified. Then we need to identify $g_1(r), g_0(r), \alpha_0, \alpha_1$ on the interval $[c_l, c_h]$. Since α_0, α_1 are constant, we only need to identify them at a single point. However, we need to be more careful about identifying g_1, g_0 at all points in the interval since they are functions of r. Below I refer to those with cutoff c_l as "low types" and those with cutoff c_h as "high types"

1. Take low types for $r < c_l$

$$\mathbb{E}[Y \mid R = r, C = c_l] = \mathbb{E}[Y(0) \mid R = r, C = c_l] = g_0(r)$$

2. Take high types for $r < c_l$

$$\mathbb{E}[Y \mid R = r, C = c_h] = \mathbb{E}[Y(0) \mid R = r, C = c_h] = g_0(r) + \alpha_0$$

   For a given $R = r$, we can subtract off the term from the first step, **which gives us point identification for α_0.**

3. Take high types for $r \in [c_l, c_h)$

$$\mathbb{E}[Y \mid R = r, c = c_h] = \mathbb{E}[Y(0) \mid R = r, c = c_h] = g_0(r) + \alpha_0$$

   Since we know α_0, $g_0(r)$ is also identified for $r \in [c_l, c_h)$. Taking limits yet again, we can identify $g_0(c_h) = \lim_{r \uparrow c_h} \mathbb{E}[Y \mid R = r, c = c_h] - \hat{\alpha_0} = g_0(c_h)$. **So $g_0(r)$ is identified on the whole interval.**

4. Take low types for $r \in [c_l, c_h)$

$$\mathbb{E}[Y \mid R = r, C = c_l] = \mathbb{E}[Y(1) \mid R = r, C = c_l] = g_1(r)$$

   Since we know g_1 is continuous, we can take $\lim_{r \uparrow c_h} \mathbb{E}[Y \mid R = r, C = c_e]$ to identify $g_1(c_h)$. **So g_1 is identified on the whole interval.**

5. Finally, take high types at $r = c_h$. For them, we know:

$$E\left[y \mid R = c_h, C = c_h\right] = \alpha_1 + g_1\left(c_h\right)$$

   Since we already identified, g_1 on our segment of interest, **this means α is identified.**

We have shown above that all the components of $\mathbb{E}[Y(1)-Y(0)|R=r]$ are identified on our interval of interest. The assumption of the continuity of $g_d(r)$ is crucial to our identification argument, especially for points right at the cutoffs. Also the implicit functional form assumption for $\mathbb{E}[Y(d)|R=r, C=c]$ is important—it specifies that g_d and α_d do not change depending on the observed value of C. This is a very important unstated assumption that makes identification fairly straightforward in our setting. Overall this is different than the normal RD setting because there are two separate cutoffs that can be realized. This is closer to a fuzzy design since compliance at the low cutoff is not complete. The upper cutoff is more like a sharp design since people above that cutoff will always be treated.

Question 5. This problem involves the data used in "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size" by Angrist and Evans (1998). The data is available on Canvas. (Note that the data has five more observations than reported in the paper. The data comes from Angrist's website, so I'm not sure what the source of this discrepancy is, but it shouldn't matter.)
For each case, the outcome is Y is worked for pay, the treatment D is more than 2 children, and the covariates X are age, age at first birth, ages of the first two children in quarters, boy 1st, boy 2nd, black, hispanic, other race. Let $Y(0), Y(1)$ denote potential outcomes, and let Z be the instrument, which varies in the parts below. Assume, as usual, that $D = \mathbb{1}[U \leq p(X,Z)]$, with p the propensity score and the usual exogeneity conditions and normalization. Let $m(d \mid u, x) \equiv \mathbb{E}[Y(d) \mid U = u, X = x]$ denote the MTR.

(a) Let Z be same sex. Estimate m for each of the following specifications. Use those estimates to estimate the ATE, ATT, ATU, and the (unconditional) LATE for same sex. Compare your results to the TSLS estimator with D as the endogenous variable, Z as excluded instrument, and X as controls (included instruments). Present your results in an easy-to-read-table. Also, construct an estimator of $m(u) \equiv \mathbb{E}[Y(1)-Y(0) \mid U = u]$, and plot it as a function of u for each specification.

i. $m(d \mid u, x) = \alpha_d + \beta_d u + \gamma_d' x$
ii. $m(d \mid u, x) = \alpha_d + \beta_d u + \gamma' x$, where γ does not depend on d
iii. $m(d \mid u, x) = \alpha_d + \beta_d u + \gamma_d' x + \delta_d' x u$
iv. $m(d \mid u, x) = \alpha_d + \beta_{d1} u + \beta_{d2} u^2 + \gamma_d' x$
v. $m(d \mid u, x) = \alpha_d + \beta_{d1} u + \beta_{d2} u^2 + \beta_{d3} u^3 + \gamma_d' x$

Estimation Steps

To estimate $m(d|u,x)$, I take the following steps. Below I use Specification 1 as an example:

1. Calculate the propensity score $p(x,z) = \mathbb{P}(D = 1|X = x, Z = z)$ using logit
2. Separately for $D = 1$ and $D = 0$, regress Y on $p(x,z)$ (which replaces u in the specification listed above) and X. This recovers the following coefficients:

$$Y_i(d) = \alpha_d^* + \beta_d^* p_i + x\gamma_d^* \text{ for } d \in \{0,1\}$$

3. Map the OLS coefficient estimates to the parameters in the given MTR specification. For example, for $\mathbb{E}[Y|D=1, P=u, X=x]$, we have:

$$\begin{aligned}
\mathbb{E}[Y \mid D=1, P=u, X=x] &= \mathbb{E}[Y(1) \mid U \leq u, X=x] \\
&= \frac{1}{u}\int_0^u \mathbb{E}[Y(1) \mid W=w, X=x]dw \\
&= \frac{1}{u}\int_0^u \left(\alpha_1 + \beta_1 w + \gamma_1' x\right) d\omega \\
&= \frac{1}{u}\left(\alpha_1 u + \frac{1}{2}\beta_1 u^2 + \gamma_1' x u\right) \\
&= \alpha_1 + \frac{1}{2}\beta_1 u + \gamma_1' x
\end{aligned}$$

So this means that $\alpha_1 = \alpha_1^*, \beta_1 = 2\beta_1^*, \gamma_1 = \gamma_1^*$. Similarly, for $\mathbb{E}[Y|D=0, P=u, X=x]$, we get:

$$\begin{aligned}
\mathbb{E}[Y \mid D=0, P=u, X=x] &= \mathbb{E}[Y(0) \mid U > u, X=x] \\
&= \frac{1}{1-u}\int_u^1 \left(\alpha_0 + \beta_0 w + \gamma_0' x\right) d\omega \\
&= \frac{1}{1-u}\left(\alpha_0 + \frac{1}{2}\beta_0 + \gamma_0' x\right) - \frac{1}{1-u}\left(\alpha_0 u + \frac{1}{2}\beta_0 u^2 + \gamma_0' x u\right) \\
&= \alpha_0 + \frac{1}{2}\beta_0 + \frac{1}{2}\beta_0 u + \gamma_0' x
\end{aligned}$$

So we have $\alpha_0 = \alpha_0^* - \beta_0^*$, $\beta_0 = 2\beta_0^*$, and $\gamma_0 = \gamma_0^*$.

4. Use the mapping of parameters to estimated coefficients to compute the MTRs. To do this, generate u, a grid of points between $[0,1]$ and use the estimated MTR parameters to calculate $m(1|u,x), m(0|u,x)$, and $m(u)$.

5. Plot the MTEs. The graphs below plot $m(u|x=\bar{x})$ as in the TA session notes.

6. Calculate all other parameters by integrating over the MTE and applying the appropriate weights. Technically, what we need to do here is calculate the unconditional MTEs $m(u)$ by calculating $m(u|x)$ for each value of $X=x$, weighting them by $\mathbb{P}(X=x)$, and summing. We could also try to do this by binning X into points of discrete support to reduce dimensionality. But essentially for each $X=x$, we would have to recompute $m_1(u|x), m_0(u|x)$ and recompute the weights for $p(x,z)$ evaluated at $X=x$. Then we could average across all the observations in our data. In the interest of time, I do not do this method on the problem set since it is computationally bulky. Instead, I only do this process for $m(u|\bar{x})$ and recover all the target parameters for X evaluated at its mean.

7. We could ideally bootstrap this whole process to get standard errors, but due to time constraints I was unable to do this here.

Parameter Mappings

Below I go through the derivation of parameter mappings for each of the specifications listed in the question. I already showed the specification 1 mappings above.

Specification 2

Since γ shouldn't vary by treatment status, I run a saturated regression here that will back out all the other values of interest. Note I don't include any interactions with the X variables The regression is:

$$Y = \alpha_0^* + \beta_0^* d_i + \beta_1^* p_i + \beta_2^* p_i d_i + x\gamma^*$$

The mapping is as follows. Define $\pi = \mathbb{P}(D=1|P=u, X=x)$. This is associated with our indicator for D in the saturated regression, so we will map any constants multiplied by this term to the β_0^*, β_2^* coefficients from the regression.

$$\begin{aligned}
\mathbb{E}[Y \mid P=u, X=x, D] &= \mathbb{E}[U(1) \mid U \leq u, X=x]\pi + \mathbb{E}[U(1) \mid U \leq u, X=x](1-\pi) \\
&= \left(\alpha_1 + \frac{1}{2}\beta_1 u + \gamma' x\right)\pi + \left(\alpha_0 + \frac{1}{2}\beta_0 u + \gamma x\right)(1-\pi) \\
&= \alpha_0 + (\alpha_1 - \alpha_0)\pi + \frac{1}{2}\beta_0 u + \frac{1}{2}(\beta_1 - \beta_0)\pi u + \gamma x \\
&\implies \alpha_0 = \alpha_0^* \qquad \alpha_1 = \beta_0^* + \alpha_0^* \qquad \beta_0 = 2\beta_1^* \qquad \beta_1 = 2(\beta_1^* + \beta_2^*) \qquad \gamma = \gamma^*
\end{aligned}$$

Specification 3

Here we have:

$$\begin{aligned}
\mathbb{E}[Y \mid D=1, P=u, X=x] &= \frac{1}{u}\int_0^u \left(\alpha_1 + \beta_1 w + x'\gamma_1 + \delta_1' x u\right) d\omega \\
&= \alpha_1 + \frac{1}{2}\beta_1 u + x'\gamma_1 + \frac{1}{2}\delta_1' x u \\
&\implies \alpha_1 = \alpha_1^*, \qquad \beta_1 = 2\beta_1^*, \qquad \gamma_1 = \gamma_1^*, \qquad \delta_1 = 2\delta_1^*
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[Y \mid D=0, P=u, X=x] &= \frac{1}{1-u}\int_u^1 \left(\alpha_0 + \beta_0 w + x'\gamma_0 + \delta_0' x u\right) d\omega \\
&= \frac{1}{1-u}\left(\alpha_0 + \frac{1}{2}\beta_0 + x'\gamma_0 + \delta_0' x\right) - \frac{1}{1-u}\left(\alpha_0 u + \frac{1}{2}\beta_0 u^2 + x'\gamma_0 + \frac{1}{2}\delta_0' x u^2\right) \\
&= \alpha_0 + \frac{1}{2}\beta_0 + \frac{1}{2}\beta_0 u + x'\gamma_0 + \frac{1}{2}x'\delta_0 + \frac{1}{2}ux'\delta_0 \\
&\implies \alpha_0 = \alpha_0^* - \beta_0^*, \qquad \beta_0 = 2\beta_0^*, \qquad \gamma_0 = \gamma_0^* - \delta_0^*, \qquad \delta_0 = 2\delta_0^*
\end{aligned}$$

Specification 4

$$\begin{aligned}
\mathbb{E}[Y \mid D=1, P=u, X=x] &= \frac{1}{u}\int_0^u \left(\alpha_1 + \beta_{11} w + \beta_{12} w^2 + \gamma_1' x\right) d\omega \\
&= \alpha_1 + \frac{1}{2}\beta_{11} u + \frac{1}{3}\beta_{12} u^2 + \gamma_1' x \\
&\implies \alpha_1 = \alpha_1^*, \qquad \beta_{11} = 2\beta_{11}^*, \qquad \beta_{12} = 3\beta_{12}^*, \qquad \gamma_1 = \gamma_1^*
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[Y \mid D=0, P=u, X=x] &= \frac{1}{1-u}\int_u^1 \left(\alpha_0 + \beta_{01} w + \beta_{02} w^2 + \gamma_0' x\right) d\omega \\
&= \frac{1}{1-u}\left(\alpha_0 + \frac{1}{2}\beta_{01} + \frac{1}{3}\beta_{02} + \gamma_0' x\right) \\
&\quad - \frac{1}{1-u}\left(\alpha_0 u + \frac{1}{2}\beta_{01} u^2 + \frac{1}{3}\beta_{02} u^3 + \gamma_0' x\right) \\
&= \frac{1}{1-u}\Big(\alpha_0(1-u) + \frac{1}{2}\beta_{01}(1-u)(1+u) \\
&\quad + \frac{1}{3}\beta_{02}(1-u)(1+u+u^2) + \gamma_0' x(1-u)\Big) \\
&= \alpha_0 + \frac{1}{2}\beta_{01} + \frac{1}{2}\beta_{01} u + \frac{1}{3}\beta_{02} + \frac{1}{3}\beta_{02} u + \frac{1}{3}\beta_{02} u^2 + \gamma_0' x \\
&\implies \alpha_0 = \alpha_0^* - \beta_{01}^*, \qquad \beta_{01} = 2(\beta_{01}^* - \beta_{02}^*), \qquad \beta_{02} = 3\beta_{02}^*, \qquad \gamma_0 = \gamma_0^*
\end{aligned}$$

Specification 5

$$\begin{aligned}
\mathbb{E}[Y \mid D=1, P=u, X=x] = \\
&= \alpha_1 + \frac{1}{2}\beta_{11}u + \frac{1}{3}\beta_{12}u^2 + \frac{1}{4}\beta_{13}u^3 + \gamma_1'x \\
&\implies \alpha_1 = \alpha_1^*, \qquad \beta_{11} = 2\beta_{11}^*, \qquad \beta_{12} = 3\beta_{12}^*, \qquad \beta_{13} = 4\beta_{13}^*, \qquad \gamma_1 = \gamma_1^*,
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[Y \mid D=0, P=u, X=x] \\
&= \alpha_0 + \frac{1}{2}\beta_{01} + \frac{1}{2}\beta_{01}u + \frac{1}{3}\beta_{02} + \frac{1}{3}\beta_{02}u + \frac{1}{3}\beta_{02}u^2 \\
&\quad + \frac{1}{4}\beta_{03} + \frac{1}{4}\beta_{03}u + \frac{1}{4}\beta_{03}u^2 + \frac{1}{4}\beta_{03}u^3 + \gamma_0'x \\
&\implies \alpha_0 = \alpha_0^* - \beta_{01}^*, \qquad \beta_{01} = 2(\beta_{01}^* - \beta_{02}^*), \qquad \beta_{02} = 3(\beta_{02}^* - \beta_{03}^*), \\
&\quad \beta_{03} = 4\beta_{03}^*, \qquad \gamma_0 = \gamma_0^*
\end{aligned}$$

Weights

I use the following weights for the MTEs to compute target parameters.

$$\begin{aligned}
\omega_{ATE} &= 1 \\
\omega_{ATT} &= \frac{\mathbb{P}(u \leq p(z))}{\mathbb{P}(D=1)} \\
\omega_{ATU} &= \frac{\mathbb{P}(u > p(z))}{\mathbb{P}(D=0)} \\
\omega_{LATE} &= \left(\frac{\mathbb{1}\left[p\left(z_0\right) < u \leq p\left(z_1\right)\right]}{p\left(z_1\right) - p\left(z_0\right)}\right)
\end{aligned}$$

Answer to Part (a)

Using the same sex instrument, Table 1 below reports estimates for each of the target parameters. Additionally, figure 1 plots the calculated MTEs and MTRs for each specification. The solid lines on the graphs are the MTEs and in specifications 1 and 3 they slope upwards, meaning that there is a larger negative effect on employment of having a child on those who are more likely to have a child (lower values of u). This is reflected in the table as well. For specifications 1 and 3, the absolute value of the ATU is smaller than the absolute value of the ATT.

Alternatively, in specification 2, there is a slightly downward sloping MTE and we see the reverse pattern, with the ATU larger in magnitude than the ATT. This suggests that there might be differences in the $Y(0), Y(1)$ distributions depending on observables X (i.e. potential selection on X).

Specifications 4 and 5 give us a quadratic and cubic fit, capturing potential nonlinearities in the MTE. Specification 4 shows positive treatment effects at low values of u and negative treatment effects in the middle of the distribution.

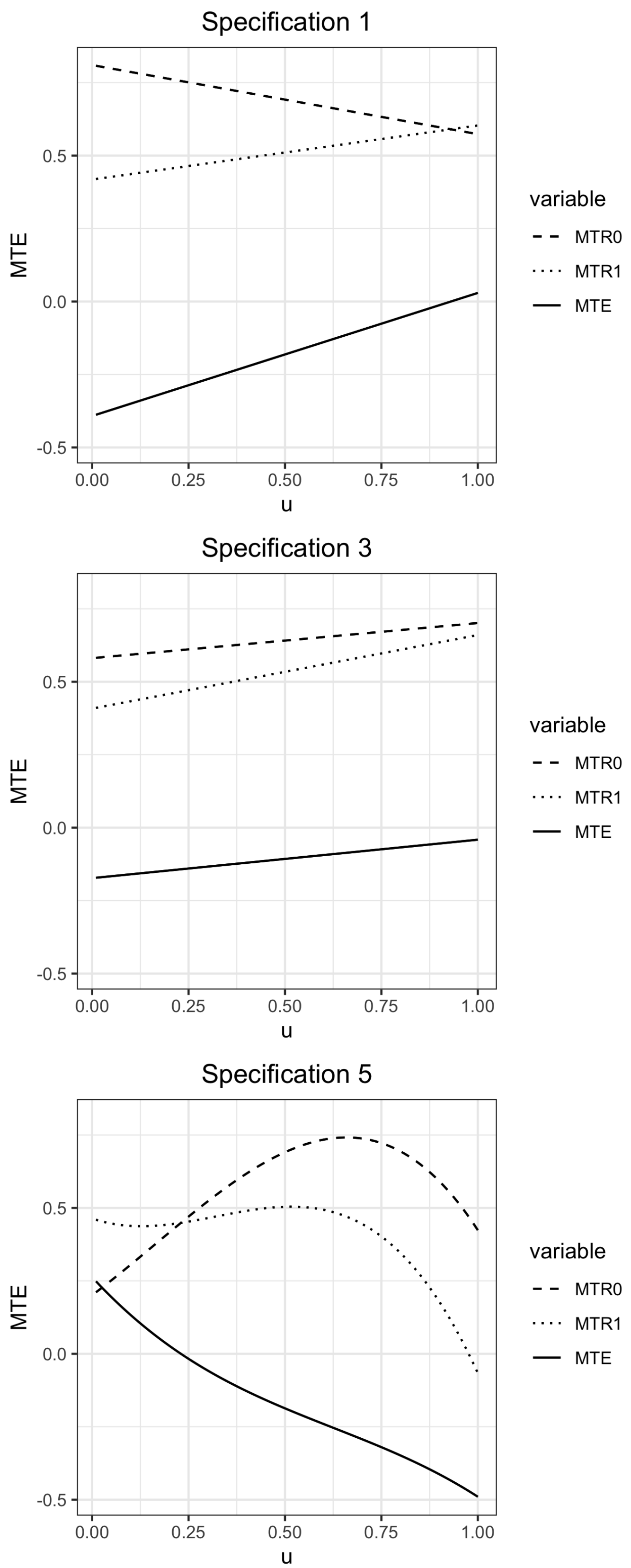
Finally, we can compare this all to a 2SLS estimation with same sex as an excluded instrument and X as controls. The estimate from that exercise is -0.118 with a standard error of 0.026.

Table 1: Treatment Effects: Same Sex Instrument

	Spec. 1	Spec. 2	Spec. 3	Spec. 4	Spec. 5
ATE	-0.1792	-0.2321	-0.1063	-0.0466	-0.1662
ATU	-0.1011	-0.2660	-0.0834	-0.1627	-0.3045
ATT	-0.2953	-0.1817	-0.1405	0.1260	0.0395
LATE	-0.2300	-0.2183	-0.1227	-0.0993	-0.1221

Note: This table reports target parameter values for all specifications in this question. These were calculated by estimating our hypothesized MTR parameters, estimating the MTR specification on a grid of points between [0,1], and weighting the MTEs at each point u by the weights given above.

Figure 1: MTEs for All Specifications: Same Sex Instrument



(b) Repeat part (a) with Z as twins. Note any additional challenges or problems that arise with this instrument.

Below we do the same exercise for the twin instrument. Again, we see upward sloping MTEs in specifications 1 and 3, a slightly downward sloping MTE in specification 2, and nonlinear fits in 4 and 5. The 2SLS estimate is -0.083 with a standard error of 0.013. We cannot rule out that the treatment effect is 0 in this case. However, looking at MTE graphs does help us say something about the relative effects for different parts of the U distribution.

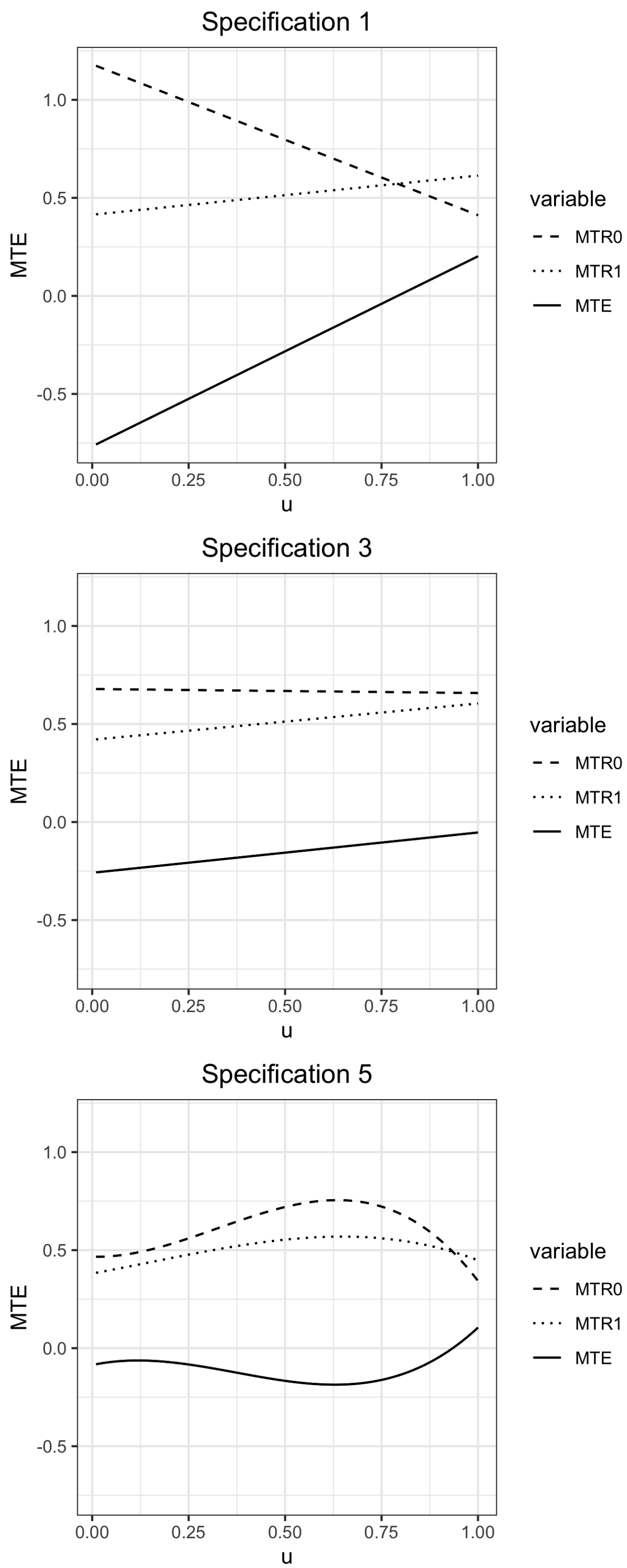
Overall, it seems that twin birth shows different results than the same sex instrument. As I discussed in the previous problem set, the twin instrument has an implicit "no defiers" assumption as opposed to the same sex instrument. The treatment effect estimates from the same sex instrument are more negative compared to the TE estimates from the twin instrument. If we believe that twin birth is a better instrument (i.e. better satisfies the necessary assumptions), then it would suggest that the people "defying" the same sex instrument are biasing the estimates downwards.

Table 2: Treatment Effects: Twin Instrument

	Spec. 1	Spec. 2	Spec. 3	Spec. 4	Spec. 5
ATE	-0.2778	-0.2279	-0.1550	0.0058	-0.1069
ATU	-0.1028	-0.2610	-0.1222	-0.1255	-0.1275
ATT	-0.5380	-0.1786	-0.2038	0.2010	-0.0762
LATE	-0.1031	-0.2487	-0.1178	-0.1220	-0.1254

Note: This table reports target parameter values for all specifications in this question. These were calculated by estimating our hypothesized MTR parameters, estimating the MTR specification on a grid of points between [0,1], and weighting the MTEs at each point u by the weights given above.

Figure 2: MTEs for All Specifications: Twins Instrument



(c) Repeat part (a) with Z as both same sex and twins.

Finally, we redo the exercise for both instruments. The tables and graphs below report the results. We see a similar pattern as in parts (a) and (b). One difference is that the nonlinear MTE in specification 5 noticeably still maintains a downward drop in the middle of the U distribution. The TSLS estimate is -0.09 with a standard error of 0.012. This is close to the part (b) results.

A note: since we have two instruments here, I calculate the LATE for each one (based on the weighting outlined above). I then combine the two LATEs by weighting them according to their relative contribution in the first stage (i.e. using the π coefficients from the first stage).

Table 3: Treatment Effects: Both Instruments

	Spec. 1	Spec. 2	Spec. 3	Spec. 4	Spec. 5
ATE	-0.2778	-0.2279	-0.1550	0.0058	-0.1069
ATU	-0.1028	-0.2610	-0.1222	-0.1255	-0.1275
ATT	-0.5380	-0.1786	-0.2038	0.2010	-0.0762
LATE	-0.1031	-0.2487	-0.1178	-0.1220	-0.1254

Note: This table reports target parameter values for all specifications in this question. These were calculated by estimating our hypothesized MTR parameters, estimating the MTR specification on a grid of points between [0,1], and weighting the MTEs at each point u by the weights given above.

Figure 3: MTEs for All Specifications: Twins Instrument

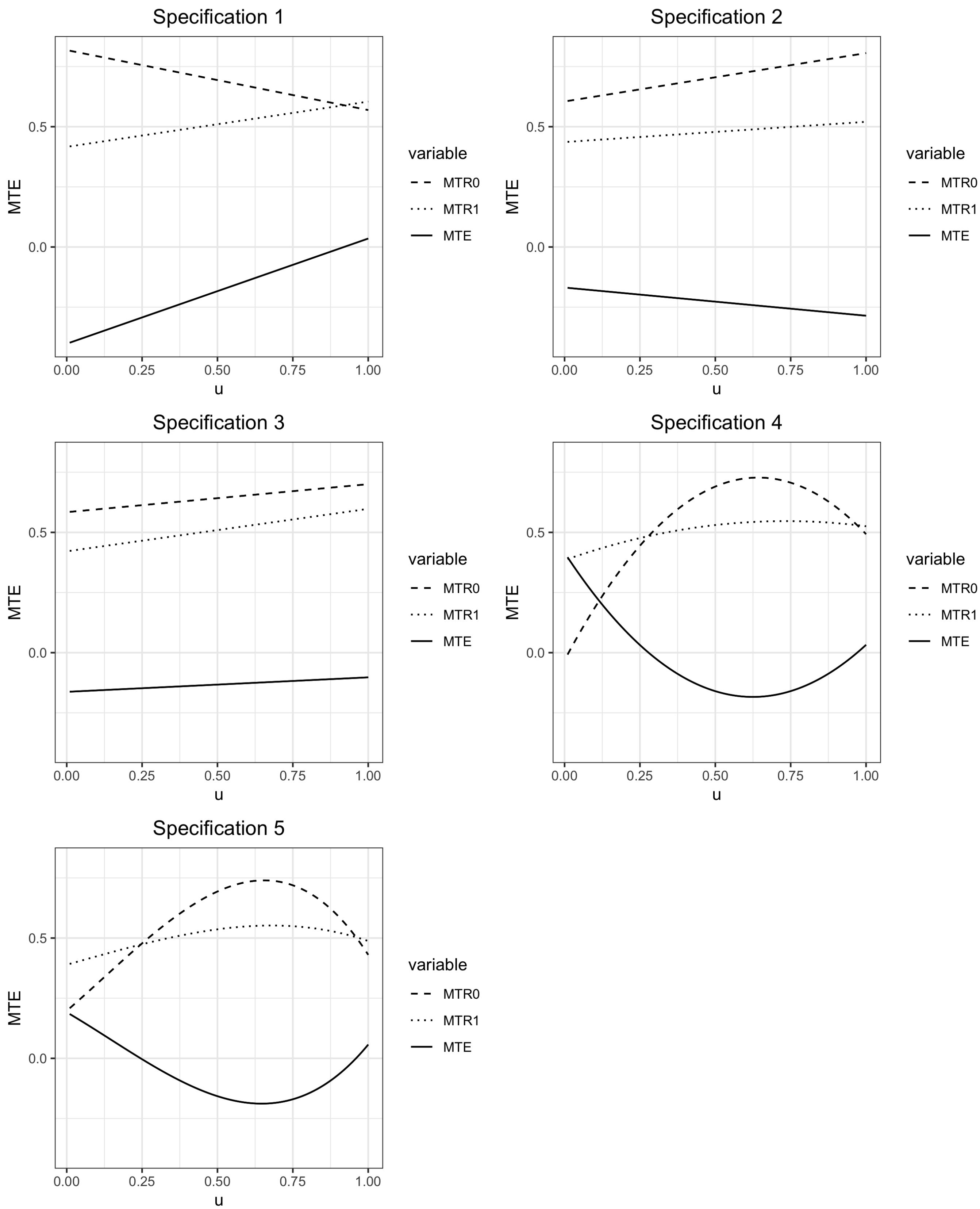


Figure 4: MTEs for All Specifications: Both Instruments

> **Question 6.** Read "Identification and Extrapolation of Causal Effects with Instrumental Variables" (2018, Annual Review of Economics) by Mogstad and Torgovitsky. Reproduce Figure 6 .

In order to reproduce Figure 6, we need to compute bounds using the optimization problem laid out in Mogstad and Torgovitsky (2018)[2]:

$$\bar{\beta}^\star = \max_{\theta \in \Theta} \sum_{d \in \{0,1\}} \sum_{k=0}^{K_d} \gamma^\star_{dk} \theta_{dk} \quad \text{subject to} \quad \sum_{d \in \{0,1\}} \sum_{k=0}^{K_d} \gamma_{sdk} \theta_{dk} = \beta_s \text{ for all } s \in \mathcal{S}$$

with the following defintions:

$$\begin{aligned}
\gamma^\star_{dk} &\equiv E\left[\int_0^1 b_{dk}(u, X)\omega^\star_d(u, X, Z)\mathrm{d}u\right] \\
\gamma_{sdk} &\equiv E\left[\int_0^1 b_{dk}(u, X)\omega_{ds}(u, X, Z)\mathrm{d}u\right] \\
\omega_{0s}(u, x, z) &\equiv s(0, x, z)\mathbb{1}[u > p(x, z)] \\
\omega_{1s}(u, x, z) &\equiv s(1, x, z)\mathbb{1}[u \leq p(x, z)]
\end{aligned}$$

Note that $\omega^*_1(u, x, z), \omega^*_0(u, x, z)$ vary by our target parameter of interest. In this problem, we want the ATT, which gives us:

$$\begin{aligned}
\omega^*_1(u, x, z) &= \frac{\mathbb{1}[u \leq p(x, z)]}{P[D = 1]} \\
\omega^*_0(u, x, z) &= -\omega^*_1(u, x, z)
\end{aligned}$$

In the definitions above, b_{dk} are basis vectors. In this problem, they can either be determined by a Bernstein polynomial of the form:

$$b^K_k(u) = \sum_{i=k}^{K} (-1)^{i-k} \left(\begin{array}{c} K \\ i \end{array}\right)\left(\begin{array}{c} i \\ k \end{array}\right) u^i$$

or nonparametrically by a constant spline of the form:

$$b^J_j(u) = \mathbb{1}\left\{u \in \mathcal{U}_j\right\} \text{ for } j = 0, \ldots, J$$

Each $\mathcal{U}_j = (u_{j-1}, u_j]$ with cutpoints u_j determined by the propensity score values. This follows from what is suggested by Mogstad, Santos and Torgovitsky (2018). In footnote 8, they explain that each $\{u_j\}_{j=0}^J$ are "the ordered unique elements of the union of $\{0, 1\}, \operatorname{supp} p(Z)$, and the discontinuity points of $\{\omega^\star_d(\cdot, z) : d \in \{0, 1\}, z \in \operatorname{supp} Z\}$." Since our $\omega^*(u, z)$ only depends on Z, and Z has discrete support, the discontinuity points on $[0, 1]$ align with the places where the propensity score jumps, giving us $J = 4$, or 5 total splines.

Calculation Steps

I take the following steps to reproduce the Figure:

[2] Lower bounds were calculated using the minimization problem.

1. Generate data using the information from section 2.4 of Mogstad and Torgovitsky (2018). This involves generating a set of $N = 50,000$ observations drawn from $U[0,1]$. Note, I only use this data to compute the IV slope and TSLS β_s and $s(d,z)$ values. Once we have these parameters, we can limit ourselves to working with the four propensity score values given in the problem (and the calculated weights associated with each one).
   (a) Calculate m_0, m_1, and m using the parameters
   $$m_0(u) = 0.9 - 1.1u + 0.3u^2 \quad \text{and} \quad m_1(u) = 0.35 - 0.3u - 0.05u^2$$
   (b) Randomly assign $\frac{1}{4}$ of the observations to each value of Z
   (c) Calculate D by replacing $D = 1$ for those with $u < p(Z = z)$ for each value of Z according to the propensity scores given:
   $$p(1) = 0.12, \quad p(2) = 0.29, \quad p(3) = 0.48, \quad \text{and} \quad p(4) = 0.78$$
   (d) Calculate $Y = m_1(u)D + m_0(u)(1 - D)$. I verify that the propensity scores match with the information from section 2.4
2. Calculate RHS of optimization constraints. The paper uses two constraints: the IV slope and the TSLS. We need to know the values of β_s for each of these and calculate them using formulas from Table 3 of the paper
   $$\beta_S^{IV} = \frac{\text{Cov}(Y,Z)}{\text{Cov}(D,Z)} = -0.277$$
   $$\beta_S^{TSLS} = e_j' \left(\Pi E\left[\widetilde{Z}\tilde{X}'\right]\right)^{-1} (\Pi E[\widetilde{Z}Y]) = -0.271$$
   where $\tilde{X} \equiv [1, D]'$ and $\widetilde{Z} \equiv [\mathbb{1}\{Z = 1\}, \mathbb{1}\{Z = 2\}, \mathbb{1}\{Z = 3\}, \mathbb{1}\{Z = 4\}]'$ and e_j' is the jth unit vector.
3. Calculate the terms in our polynomial. We need to get each of the γ_{sdk} terms, so we need to calculate the functions $s(0,x,z), s(1,x,z)$. We can again look in the specifications from table 3 to calculate this for the IV slope and the TSLS.
   $$s(d,z)^{IV} = \frac{z - E[Z]}{\text{Cov}(D,Z)}$$
   $$s(d,z)^{TSLS} = e_j' \left(\Pi E\left[\widetilde{Z}\tilde{X}'\right]\right)^{-1} \Pi\widetilde{Z}$$
4. From the Mogstad, Santos, and Torgovitsky (2018) supplement, we can write each γ_{sdk} as:
   $$\gamma_{sdk} \equiv E\left[\int_0^1 b_{1k}(u,Z)\omega_{1s}(u,Z)du\right] = E\left[s(1,Z)\int_0^{p(Z)} b_{1k}(u,Z)du\right]$$
   This makes it relatively easier to compute each of the γ_s values.
5. Find the γ_{dk}^* using the ω^* weights for the ATT parameter.
6. Set up the linear programming problem
   (a) Pass the Gurobi solver the objective function (a vector with entries $[\gamma_{0k}^* \quad \gamma_{1k}^*]$)
   (b) Pass the solver IV and TSLS constraints, which should be equal to β_s^{IV} and β_s^{TSLS}, respectively. These are also similar vectors of γ_{sdk} values.

(c) Create shape constraints as outlined in the Mogstad, Santos, Torgovitsky (2018) supplement such that $\theta_{0,d} \geq \theta_{1,d} \geq \cdots \geq \theta_{K,d}$

(d) Set bounds so that $\theta_k \in [0, 1]$ for all k

Figure 5: Upper and Lower Bounds on the ATT